From prediction to prevention: Machine learning revolutionizes hepatocellular carcinoma recurrence monitoring

Mariana Michelle Ramírez-Mejía, Nahum Méndez-Sánchez

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report’s scientific quality classification
Grade A (Excellent): 0
Grade B (Very good): 0
Grade C (Good): C, C
Grade D (Fair): 0
Grade E (Poor): E

P-Reviewer: Jin C, China; Yu H, China

Received: November 13, 2023
Peer-review started: November 13, 2023
First decision: December 5, 2023
Revised: December 12, 2023
Accepted: January 22, 2024
Article in press: January 22, 2024
Published online: February 21, 2024

Abstract

In this editorial, we comment on the article by Zhang et al entitled Development of a machine learning-based model for predicting the risk of early postoperative recurrence of hepatocellular carcinoma. Hepatocellular carcinoma (HCC), which is characterized by high incidence and mortality rates, remains a major global health challenge primarily due to the critical issue of postoperative recurrence. Early recurrence, defined as recurrence that occurs within 2 years posttreatment, is linked to the hidden spread of the primary tumor and significantly impacts patient survival. Traditional predictive factors, including both patient- and treatment-related factors, have limited predictive ability with respect to HCC recurrence. The study by Zhang et al demonstrated the use of a groundbreaking preoperative prediction model for early postoperative HCC recurrence. Challenges persist, including sample size constraints, issues with handling data, and the need for further validation and interpretability. This study emphasizes the need for collaborative efforts, multicenter studies and comparative analyses to validate and refine the model. Overcoming these challenges and exploring innovative approaches, such as multi-omics integration, will enhance personalized oncology care. This study marks a significant stride toward precise, efficient, and personalized oncology practices, thus offering hope for improved patient outcomes in the field of HCC treatment.

Key Words: Hepatocellular carcinoma; Early recurrence; Machine learning; XGBoost model; Predictive precision medicine; Clinical utility; Personalized interventions
INTRODUCTION

In this editorial, we comment on the article by Zhang et al [1] entitled Development of a machine learning-based model for predicting the risk of early postoperative recurrence of hepatocellular carcinoma. Hepatocellular carcinoma (HCC) is the most common form of primary liver cancer and is considered a major global health challenge due to its high incidence and mortality rates [2,3]. Despite advances in medical and surgical interventions, recurrence remains a critical problem affecting the long-term survival of patients with HCC [4,5]. Recurrence of HCC within the 2 years posttreatment is categorized as early recurrence. Early recurrence typically occurs due to the hidden spread of the primary tumor within the liver, and its incidence is correlated with the tumor’s size and extent. On the other hand, recurrence after 2 years posttreatment is categorized as late recurrence. Late recurrence is associated with de novo HCC, indicating the development of new cancerous growth independent of the original tumor [6]. Several predictive factors associated with recurrence have been recognized. Factors contributing to early recurrence include patient-related aspects such as age, the presence of underlying health conditions, liver function, viral load, the presence and activity of hepatitis, metabolic dysfunction-associated fatty liver disease, alcoholic liver disease and other etiologies, and the existence and activity of liver cirrhosis. Additionally, treatment-related factors, including the type of treatment employed, surgical margins, and specifics of the resection, also play a crucial role in predicting early recurrence [7]. The intricate nature of liver cancer, coupled with the diverse factors influencing recurrence, makes it challenging to provide an accurate prognosis [8,9]. Due to the constantly evolving landscape of HCC research, the quest for methods for predicting early recurrence has undergone a remarkable transformation in recent decades. Initially, researchers focused on deciphering the morphological characteristics of tumors as a basis for predictions [10]. Factors such as vascular invasion, tumor multiplicity and large tumor size have emerged as fundamental, although somewhat rudimentary, indicators that establish the basis for understanding the complexities of HCC recurrence [11,12]. Nevertheless, the paradigm shifted with the arrival of molecular analysis. Elevated alpha-fetoprotein (AFP) levels emerged as one of the first markers used for HCC prediction, offering insight into the intricate molecular landscape of this aggressive cancer [13,14]. Despite these advances, the multifaceted nature of HCC recurrence requires more nuanced and sophisticated approaches. Researchers and clinicians alike recognize the limitations of relying solely on morphologic and molecular analyses [15]. The quest for increasing predictive accuracy has led the scientific community to explore uncharted territory and harness the transformative power of technology, especially in the fields of imaging, genetics and computational sciences [16].

The evolution of technology has emerged as a pivotal catalyst, propelling HCC research into an era of unprecedented possibilities. Cutting-edge imaging techniques, coupled with advancements in genetic profiling, provide researchers with a comprehensive understanding of the tumor microenvironment [17-20]. These insights, combined with the computational progress of modern data analysis, paved the way for a new generation of predictive models. These models transcended the limitations of traditional analyses, offering a more nuanced and accurate glimpse into the future course of HCC [21-24].

MACHINE LEARNING IN HCC RESEARCH

The exponential growth of computational power has heralded a new era in HCC research, where machine learning algorithms have emerged as valuable tools in handling vast datasets and deciphering complex patterns [25,26]. This convergence of computational capabilities and healthcare needs represents a significant paradigm shift, transforming the landscape of HCC research. The integration of machine learning into the study of HCC offers a multitude of benefits and promises to address long-standing challenges in this field [27,28].

In a retrospective study, Zhang et al [1], harnessed the potential of supervised machine learning to develop a state-of-the-art preoperative prediction model for early postoperative HCC recurrence. Leveraging readily available clinical and imaging data, the team built six different risk prediction models, using ensemble learning, linear and neural network models, each meticulously designed to identify patients at high risk of recurrence. The study methodology consisted of analyzing the demographic and clinical data of 371 patients with HCC, excluding cases with incomplete data or previous
neoadjuvant treatments. Using machine learning algorithms, the researchers identified eight key variables to predict early HCC recurrence: Age, intra-tumoral arteries, AFP, blood glucose, number of tumors, glucose-to-lymphocyte ratio (GLR), liver cirrhosis, and platelet count. These variables formed the basis for the construction of six different prediction models, of which the XGBoost model proved to be the most robust (Figure 1). The XGBoost model, outperformed its peers, showing unmatched performance on several metrics. In the training dataset, the model achieved an impressive area under the receiver operating characteristic curve (AUROC) of 0.993, proof of its accuracy. Even in the validation and test data sets, the XGBoost model maintained its excellence, with AUROC values of 0.734 and 0.706, respectively. Calibration curve analysis underscored the reliability of the model, confirming its alignment with real-world results. Furthermore, decision curve analysis highlighted the clinical utility of the XGBoost model, highlighting its potential to guide surgical strategies and usher in an era of individualized postoperative medicine. By employing the SHAP (SHapley Additive exPlanations) package, the study provided a detailed interpretation of the model results, unraveling the intricate relationships between variables. Preoperative glycemia emerged as a key factor, in line with previous research highlighting its role in HCC progression. The predictive power of the model was further demonstrated using an online calculator, designed to assist physicians in their daily practice. This user-friendly tool represents a major breakthrough, as it ensures the seamless integration of predictive analytics into clinical decision making.

The relevance of this study goes far beyond conventional medical research. Machine learning algorithms, used to decipher the intricate web of preoperative variables, have ushered in an era where predictive precision medicine reigns supreme. The identification of these eight key variables is a pivotal moment, providing physicians with unprecedented insight into the intricate factors that determine early postoperative recurrence.

Several research projects have been conducted to explore the use of machine learning in predicting HCC recurrence. Kucukkaya et al.[29] developed a predictive model based on the analysis of pre-treatment magnetic resonance imaging using the VGG16 and XGBoost machine learning models. This model aimed to predict recurrence in six different time intervals, ranging from 1 year to 6 years, and demonstrated performance with AUROC values between 0.71 and 0.85. In another study, Zeng et al.[25] compared the performance of random survival forest (RSF) models with Cox proportional hazard (CPH) models in predicting early recurrence using clinical features of the participants. In training and internal and external validation cohorts, the C-index of the RSF model was 0.725, 0.762, and 0.747, respectively. Although both studies highlighted the utility of machine learning, neither included the analysis of clinical and imaging variables, assuming a linear interaction of predictors for HCC recurrence. In this context, Zhang et al.[1] proposed a solution in their
study, addressing the need to include clinical and imaging variables in the analysis. Their approach seeks to overcome the limitation of assuming linear interactions among predictors of HCC recurrence.

---

**CHALLENGES AND FUTURE DIRECTIONS**

This represents a significant advancement in early postoperative HCC recurrence prediction. Future research should focus on overcoming challenges related to sample size, data handling, validation, and interpretability. The authors acknowledge these limitations, emphasizing the need for future research endeavors to validate and refine the model further. These findings call for additional research, urging the scientific community to collaborate, expand sample sizes, and conduct multicenter studies. Comparative analyses with existing prediction models are crucial for ensuring the reliability and applicability of this innovative approach [30, 31].

---

**CONCLUSION**

The development of accurate, interpretable, and widely applicable prediction models for early postoperative HCC recurrence represents a significant advance in personalized medicine. Addressing the challenges associated with data quality and model interpretability while exploring innovative approaches, such as multi-omics integration and continuous model refinement, will pave the way for improved patient outcomes and healthcare practices in the field of HCC treatment. Through collaborative efforts, continued research and the use of patient-centered approaches, the field of oncology can continue its journey toward more precise, efficient and personalized oncology care.

---

**FOOTNOTES**

**Author contributions:** Méndez-Sánchez N and Ramirez-Mejia MM contributed to this paper; Méndez-Sánchez N designed the overall concept and outline of the manuscript; Ramirez-Mejia MM contributed to the discussion and design of the manuscript; Méndez-Sánchez N and Ramirez-Mejia MM contributed to the writing and editing of the manuscript, the illustrations, and the review of the literature.

**Conflict-of-interest statement:** All the authors declare that they have no conflicts of interest related to the manuscript.

**Open-Access:** This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: https://creativecommons.org/Licenses/by-nc/4.0/

**Country/Territory of origin:** Mexico

**ORCID number:** Mariana Michelle Ramírez-Mejía 0000-0005-6279-1527; Nahum Méndez-Sánchez 0000-0001-5257-8048.

**S-Editor:** Lin C
**L-Editor:** Filipodia
**P-Editor:** Chen YX

---

**REFERENCES**


